Journal of Digital Imaging

Effect of Automated Image Registration on Radiologist Interpretation

Bradley J. Erickson,[1] Jayawant Mandrekar,[2] Liqin Wang,[3] Julia W. Patriarche,[1] Brian J. Bartholmai,[1] Christropher P. Wood,[1] E. Paul Lindell,[1] Anne-Marie Sykes,[1] Gordon F. Harms,[1] Rebecca M. Lindell,[1] and Katherine Andirole[4]

In this study, we present preliminary data on the effect of automated 3D image alignment on the time to arrive at a decision about an imaging finding, the agreement of multiple of multiple observers, the prevalence of comparison examinations, and technical success rates for the image alignment algorithm. We found that automated image alignment reduced the average time to make a decision by 25% for cases where the structures are rigid, and when the scanning protocol is similar. For cases where these are not true, there is little or no benefit. In our practice, 54% of cases had prior examinations that could be automatically aligned. The overall benefit seen in our department for highly similar exams might be 20% for neuro and 10% for body; the benefit seen in other practices is likely to vary based on scanning practices and prevalence of prior examinations.

KEY WORDS: Image registration, image alignment, practice efficiency, TRIP

INTRODUCTION

Computer algorithms for automated spatial alignment (also known as image registration) of separate 3-dimensional images have been described for more than 2 decades.[1–9] They are of value for many image processing purposes. The evaluation of these algorithms in clinical practice has been very limited.[10,11] Perhaps as a consequence, these algorithms have not been available on PACS workstations or used routinely in the interpretation process by radiologists and, therefore, have had little impact on the practice of medical image interpretation. We studied the effect of image registration on interpretation of typical radiology studies by a radiologist in simulated reading environment. We hypothesized that if images from a prior examination with similar characteristics are aligned to the examination to be interpreted, it will reduce the amount of time a radiologist spends looking at the images and will increase agreement between observers. If this case is proven, wider use of image registration in radiology may be of value, and integration of registration tools onto PACS workstations could increase radiologist productivity and consistency. We should note that in the literature, this process is also referred to as "registration" or "rigid registration" and we will use those terms interchangeably with "alignment".

METHODS

Patient/Exam Selection

After IRB approval, we collected sequential computed tomography (CT) or magnetic resonance imaging (MRI) cases from our body and

[1]From the Department Radiology, Mayo Clinic, 200 First St SW, Rochester, MN, 55905, USA.

[2]From the Department Biostatistics, Mayo Clinic, 200 First St SW, Rochester, MN, USA.

[3]From the Department Information Services, Mayo Clinic, 200 First St SW, Rochester, MN, USA.

[4]From the Department Radiology, Brigham and Women's Hospital, Boston, MA, USA.

Correspondence to: Bradley J. Erickson, Department Radiology, Mayo Clinic, 200 First St SW, Rochester, MN, 55905, USA; tel: +1-507-2846238; fax: +1-507-2842405; e-mail: bje@mayo.edu

doi: 10.1007/s10278-007-9023-x

neuroradiology practices which are considered the "current" examination for the purposes of this study, and "neuro" refers to images of the head, brain, neck, and spine; "body" refers to all other body parts. For each of the four examination groups (neuro CT, neuro MR, body CT, and body MR), we recorded if there were prior MR or CT exams of the same body part as this "current" examination (cross-modality matches were included, e.g., prior CT of the left knee would be considered a match for current MR examination of the left knee). We collected the first 20 cases with such a prior matching examination for body CT and MR as well as 20 for neuro CT and MRI which had at least one positive finding. There was one neuro MR case that had a prior spine CT, and two CT cases that had prior MR examinations; no body CTs had an MR as the prior examination, nor were there body MRs that had prior CT examinations. The exact mix of CT and MR in each group and the organ/disease being imaged thus reflected our practice mix (about 80% of neuro is MRI, 80% of body is CT). We also collected the fraction of cases that had priors.

We transferred these examinations, and the most recent MR or CT comparison examination to a research PACS workstation that is identical to our clinical workstations. The studies were deidentified according to HIPAA regulations and patient identification on each pair of examinations (current + prior) was replaced with study identifiers. For each pair of examinations, a radiologist (BJE) identified a single reference finding. Examples of the reference finding included "left kidney abscess", "cervical spondylosis", and "left frontal glioma". The reference finding had to be present on the prior examination and be deemed of clinical importance.

Examination Alignment

Once the studies had been retrieved, a radiologist (BJE) reviewed each of the series of the current examination and found the most appropriate matching series from the prior examination. Criteria for matching included similar imaging plane and similar contrast properties. In some cases, a given series from the prior examination was used more than once (e.g., if the prior exam consisted only of a contrast-enhanced CT in the axial plane, all axial MR sequences had this same CT series considered as matching). In a few cases, there was no equivalent series, in which case a series of black images was created (e.g., a sagittal MR sequence in which the prior exam only consisting of axial CT images).

For each current exam series, the matching prior exam series was spatially registered using an application we developed which was based on the Insight Toolkit (http://ITK.org) multiresolution mutual information algorithm. Some of the implementation specific parameters include: five multiresolution levels with "shrink factors" of 8, 8, and 1 (x, y, z) as described in the ITK documentation. We used 100 iterations at each level, except the final level, where we used 200 iterations. We used the Mattes' Mutual Information metric[12], a Quaterion rigid transform, B-spline interpolation (see Appendix A for a description of these steps). RegularStepGradientDescentOptimizer with a large initial search space (15 units) was used to register images with a large translation or rotation compared to the current/current image set. If the registration fails, the program automatically tries alternative values until the right initial search parameter is found or the number of retries exceeds five. Instead of advancing parameters in the direction of the gradient with bipartition scheme (which is the default in the ITK code), we used a smaller scale, which is 90% of the previous values. We split double echo MRI series into a proton density (PD) and a T2 series, and used the T2 series to do the registration, and then applied that transform to the PD series as well.

In some cases, more significant manipulation was required before alignment could be applied. For instance, it is common practice to acquire multiple groups of obliquely aligned images through each lumbar interspace when performing MRI. However, with CT, it is more customary to acquire one large volume, and then create subvolumes angled with each disk space. Therefore, with lumbar spine CT exams, sometimes, each subvolume is a series, while in other cases all of the subvolumes were placed in one series. Therefore, it was necessary to review each group, break each subvolume into its own series, and align that disk space subvolume with the prior disk space subvolume.

After the series were registered, a radiologist (BJE) reviewed each series and noted those that did not align successfully, and used the interactive volume matching module in the Analyze™ software (Version 6.0, Mayo Clinic, Rochester, MN, USA) to achieve satisfactory alignment. In a few cases, satisfactory alignment could not be achieved due to poor image quality (e.g., significant patient motion artifact) or significant change in anatomy between studies. In these cases, as in those cases with no corresponding prior series, a series of black images was created and used as a placeholder to integrate with the hanging protocol feature of the display application.

After the series were aligned, the slices from the registered comparison data set were exported as a DICOM series, with appropriate spatial header elements to ensure proper alignment on a PACS workstation. For the unregistered control data sets, the series number was altered so that matching series were correct, but no spatial alignment was performed.

At this point, three sets of examinations were present: the current examination, the original prior examination with the original image data (but with series numbers changed to match the current examination), and the prior examination with the registered image data and matching series numbers. We felt it was critical to alter the series numbers for the prior studies to make the arrangement similar, and truly focus on the value of registration alone. This also necessitated the "black" series as placeholders. We also altered the patient names such that the current and registered prior exams were called "Regnn" where nn was the case number, and the other pair of current and prior exams had the patient name "Unregnn".

These four examinations were transferred to a sienet V42 (Siemens Medical Systems, Erlangen, Germany) Picture Archive and Communications Systems (PACS) display station in our Radiology Informatics Laboratory. The examinations were again visually reviewed to assure that all examinations had transferred properly and that corresponding series were displayed in a matching fashion. We used a hanging protocol in which the current examination was displayed across the top of 4 DICOM calibrated 21″ diagnostic gray scale monitors, and the prior examination was displayed across the bottom, with seven large display areas for the first seven series of each examination, and "thumbnail" series on the right-most monitor that could be dragged and dropped onto any of the other display areas (see Fig. 1).

Interpretation

As noted above, for each exam, a single reference finding was selected as the finding the radiologist should focus on. For that finding, a rating scale of −3 to +3 was used, where −3 meant marked worsening, 0 meant no change, and +3 meant marked improvement. In some cases, a tumor had been removed or resolved (and there was no other candidate reference finding) and so the scale was −3 (significant tumor remaining) to 0 (no tumor remaining). We used a single reference

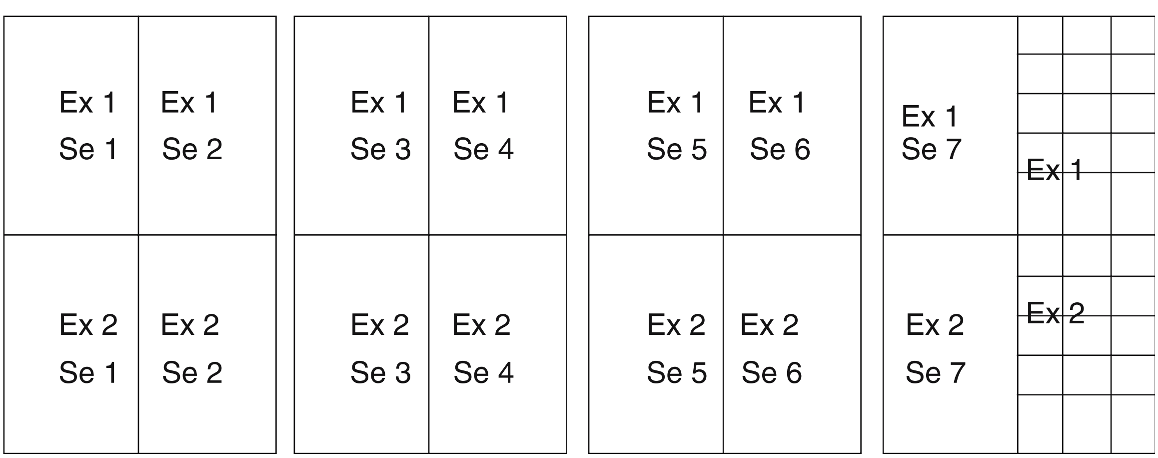

Fig 1. Arrangement of images across the four portrait-oriented PACS monitors. If there were more than seven series, the additional series would be displayed at the right side of the right-most monitor, and could be "dragged" to the desired display location. Ex 1 was always the "current" exam and Ex 2 was always the prior exam; the series numbers for the current examination were as scanned, and the prior examination was reordered to match the current.

finding to avoid problems of determining if radiologists were reporting the same or different findings, and to avoid the vagaries of human language.

Three staff neuroradiologists and four staff body radiologists were recruited to participate in reviewing these examinations in two crossover interpretation sessions which were separated by at least 4 weeks. The three neuroradiologists reviewed the 20 neuro cases and the four body radiologists reviewed the 20 body cases. During a reading session, the radiologist would alternately review registered and unregistered cases. In one session, all cases were reviewed—half with the prior examination registered, and half with the prior examination unregistered (no alteration other than removing identifiers and remapping the series number). In the second session, all cases were reviewed again, but with the opposite set of registered/unregistered prior exams. No attempt was made to blind reviewers as to whether an examination had been registered or not because the presence of registration was quite obvious: unless the prior study was acquired exactly aligned with the current (this never happened and is probably not possible—even acquiring new aligned with prior is hard) the "edges" where the prior did not align was obvious, and therefore could not be blinded.

A proctor (LW) assisted the radiologists for all sessions, and would tell them the reference finding as they opened the case. Once all the images were loaded, the radiologist would say "go" and begin reviewing the images to assess that reference finding. At that time, the proctor began a timer that was implemented into the Excel (Microsoft Corporation, Redmond, WA, USA) spreadsheet used for collecting the ratings. When the radiologist had arrived at a decision, she/he would say "stop" and then say the rating. The proctor would stop the timer (and the time was automatically recorded) and enter the rating (−3 to +3). Radiologists were neither allowed to review their ratings nor the ratings of others.

Data Analysis

Basic statistical analyses were conducted on the recorded findings and timing data, as well as the failure rate of the alignment algorithm, the number of series in an examination that was similar (e.g., same contrast properties and plane of acquisition).

We included all examinations performed until recruitment was filled, and did not attempt to alter practice behavior. We noted that in some cases, patients were scanned very much like the prior examination, while in other cases, there was a significant change in scanning methods. We defined "similar" examinations as those in which there was at most one series which was significantly different from the prior (and also the majority of images such as in a two series CT with a scout series) and "dissimilar" as those where at least 50% of the series or images were different. Differences include a different scanning plane or substantially different contrast properties.

Statistical Methods

Time spent reviewing registered versus unregistered images was compared using Wilcoxon signed rank test due to the smaller sample sizes and non-Gaussian distribution of the data. This was done by combining all the images from the reviewers as a group (Neuro or Body) and also separately for each of the reviewers. Such comparisons were also made for each reviewer by stratifying on the similarity or dissimilarity of exams.

The agreement among the radiologists was assessed using intraclass correlation[13] and appropriate 95% confidence intervals were estimated. This was done to assess overall performance as well as by stratifying within similarity status.

RESULTS

Impact on Radiologist "Decision Time"

There was a statistically significant difference in the mean "decision time" for registered versus unregistered comparison studies for neuroradiologists (Fig. 2). We found that for similar examinations in neuroradiology, the decision time was reduced from a mean of 112 to 68 s and for body cases, it was reduced from 101 to 82 s. The benefit of alignment was greater for those examinations where the series were similar. In examinations where similarity was low (dissimilar examinations), the value of aligning

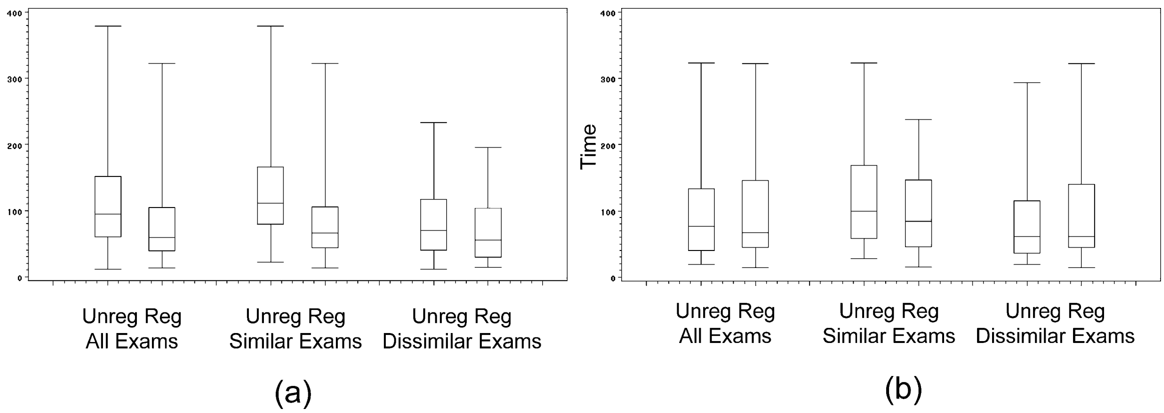

Fig 2. Decision time for registered versus unregistered examinations, and for examinations with high similarity of series, and for low similarity of series. (a) Shows results for neuroradiology and (b) is for body images. Unreg = unregistered images, Reg = registered images. Similar examinations are those where the majority of series had the same contrast properties and scan plane as the current examination.

prior examination appears low and may even slow decision time.

In our practice, 52% of neuro MRI examinations had relevant prior examinations, 62% of neuro CT examinations had relevant priors, 54% of body CT had relevant priors, but only 22% of body MRI had relevant priors. Given the overall volumes of these examinations in our practice, 54% of neuro and body CT and MR examinations have relevant prior examinations that could be registered to the current examination.

Given the reduction in decision times and the frequency of examinations that had relevant priors in our practice, one could expect an overall reduction in decision time of 20% for neuro examinations that were highly similar and 10% for body examinations that were highly similar.

Impact on Radiologist Agreement

Table 1 shows the intraclass correlation for registered versus unregistered examinations. Only neuro images with high similarity showed a statistically significant difference ($p<0.02$), but all examinations as well as both similar and dissimilar subgroups showed a trend for better agreement with registration.

Algorithm Performance

The adapted ITK algorithm for image registration successfully aligned 157 of 168 series, the remainder requiring manual alignment[8] or simply could not be aligned.[3] Manual alignment typically was accomplished by adjusting the range of

Table 1. Intraclass Coefficients Comparing Registered Versus Unregistered Images

	Number of Cases	Registered	Unregistered	*p*-Value
Neuro (all)	20	0.87	0.76	NS
Neuro (similar exams)	12	0.84	0.49	<0.02
Neuro (dissimilar exams)	3	0.91	0.87	NS
Body (all)	20	0.93	0.88	NS
Body (similar)	7	0.89	0.84	NS
Body (dissimilar)	5	0.95	0.88	NS

In all cases, there was a trend for better agreement with registered images, though only neuro (brain, head, neck, and spine) exams with similar scanning protocols showed statistical significance. We defined "similar" examinations as those in which there was at most one series which was significantly different from the prior (and also the majority of images such as in a two series CT with a scout series) and "dissimilar" as those where at least 50% of the series or images were different. Differences include a different scanning plane or substantially different contrast properties.

intensities included in the search set, or by manually aligning images to an approximate solution, allowing the computer to fine-tune the manual estimate. This typically required 15–20 s. A significant factor in the success rate of alignment was the degree of similarity between the series, which is not surprising: all of the series that had high similarity matched.

The median computation time to register one series with another was 64 s on a 2 Ghz P4 (Intel Corporation, Santa Clara, CA, USA) with 1GB RAM. This time could be reduced, but at the cost of a lower success rate for correct alignment. We could not increase the success rate significantly without huge increases in computation time. One of the components of the search is that we perform an inverse search. This means that we first do the forward registration of the prior examination onto the new examination and then compute the inverse—the new examination is registered to the old. We then require that geometrical transformation of the old examination onto the new be the inverse of the new examination onto the old. When this test fails, it suggests the solution may not be correct; adjusting the starting parameters will usually result in a solution that is correct. If no correct inverse solution is found after three tries, the algorithm "fails" but still creates a match volume in case it is "good enough".

DISCUSSION

Image registration does appear to improve rater agreement and improve efficiency. This study did not seek a gold standard of a known diagnosis with proof of the degree of change, though in most cases, the disease was known. The degree of change, particularly for nonmeasurable but evaluable disease, may not have a gold standard. We used agreement of reviewers as a surrogate for truth. We feel it is reasonable to assume that agreement of multiple experts on the magnitude of change is likely a good indicator of the truth, and hence that was used in this study. A consensus panel could also have been used but would have increased the cost of the study, and would only reflect agreement of our expert reviewers.

For many cases, changes are dramatic, and image registration is not needed. However, for subtle changes, the technique can be quite valuable. Figure 3 shows an example of an oligodendroglioma that has enlarged slightly between the two examinations. While some observers felt there was some progression, they were not unanimous in that opinion and all were equivocal without registration. With registration, all were certain that progression was present. We provide a subtraction image for this manuscript (this was not provided for this study) that helps demonstrate that the readers were likely more correct with alignment. It is in these subtle cases that the technique is most valuable.

To estimate the final expected value of image registration to practice efficiency, it was necessary to know the fraction of examinations that had prior examinations that might be registered. As noted above, in our practice, 52% of neuro MRI exams had relevant prior examinations, 62% of neuro CT examinations had relevant priors, 54% of body CT had relevant priors, but only 22% of body MRI had relevant priors. Given the overall volumes of these examinations in our practice, 54% of neuro and body CT and MR examinations have relevant prior examinations that could be registered to the current examination. This percentage will vary among practices, and this will affect the total benefit one might expect from image registration. Given that the median decision time for neuroradiology MR and CT images is approximately 25% faster with registration, and given that about half of cases have priors to register, the net benefit in terms of reduced interpretation time could be as large as 12.5%, as well as the benefit of greater interobserver agreement, suggesting higher accuracy. An important assumption in this is that decision time about change is the dominant component of interpretation—we will discuss this assumption later. The benefit to body imaging appears smaller—perhaps 3%.

Some might be disappointed in the rather small gain found in this study. Indeed, in a prior study of head CTs, Schellingerhout et al.[10] found a much more dramatic gain of about 65%. This is probably a reflection of the very different examination selection method used. The study of Schellingerhout et al. selected for complex examinations that were acquired in very similar ways. This is the type of case where the greatest gains were seen in our study. Similarly, if the prior examination was acquired in such a different

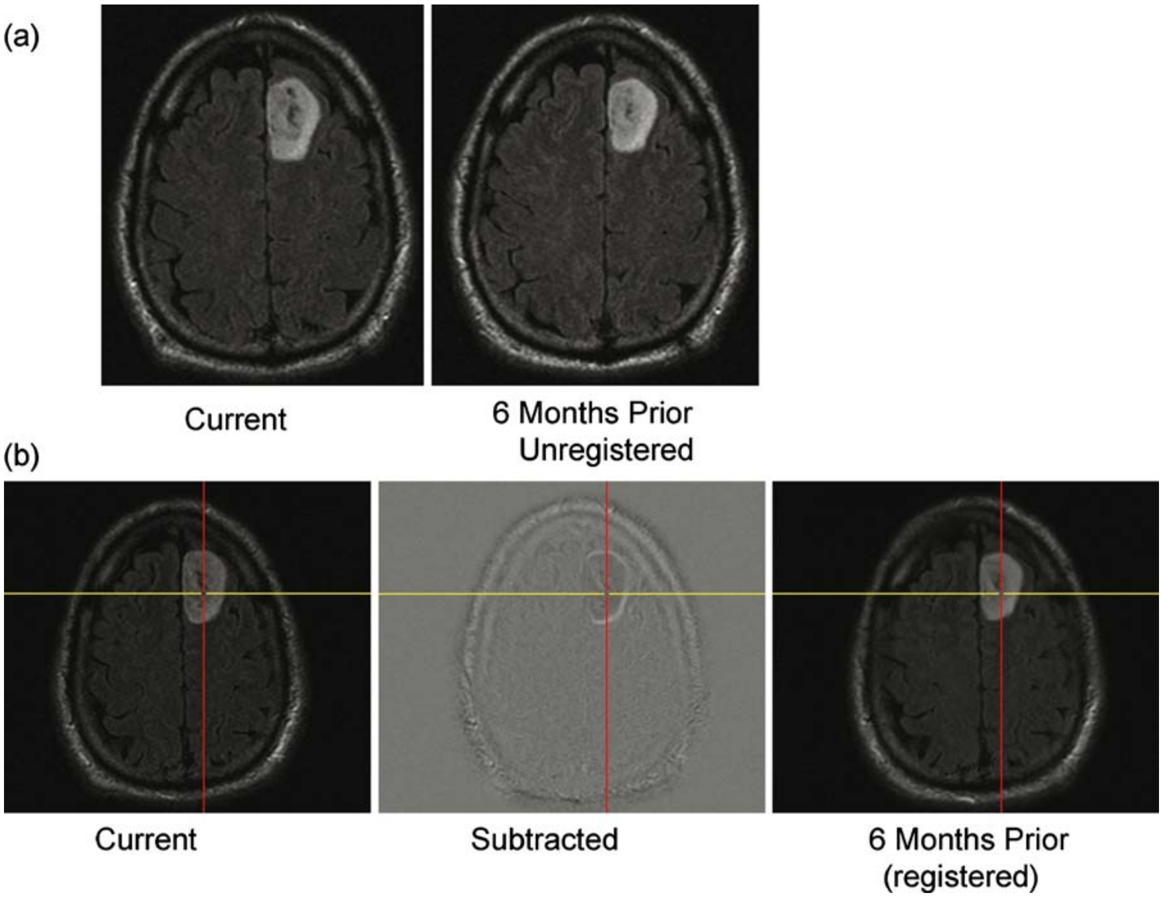

Fig 3. a Shows a single FLAIR image from the current and from a 6 months previous examination in a patient with an oligodendroglioma. Two of three observers indicated probable increase in size over the interval. b Shows the registered images as well as a subtraction image (subtraction was not provided for this study) which clearly demonstrates interval progression.

fashion that there were significant artifacts that rendered the comparison image nearly useless or confusing, there would be little advantage or even a disadvantage. This problem can be addressed by standardization of scanning methods (often driven by hanging protocols which also benefit from protocol standardization) and from scanning methods that produce near isotropic voxel proportions.

We note in this study that the task in the study was not the full interpretation task that occurs in reading a study. The finding of interest was known—only assessment of the change. It is likely that the search task of detecting the finding will not be substantially altered because that occurs primarily on the current examination. However, many observer studies have shown that trained radiologists efficiently detect findings, and spend most of their time deciding if a finding is significant, and assessing its nature. Alignment would likely help with these tasks: if a possible finding was present before and appears the same, it is probably not significant, or at least one could likely use the prior report as a guide for interpretation.

Tan et. al.[14] have reported on the use of image registration and subtraction to improve the interobserver agreement for changes in T2 lesions in patients with multiple sclerosis. They found good agreement between observers for current and enlarging lesions, but poor agreement for resolving lesions. In a similar study using FLAIR images in multiple sclerosis (MS), they found that registration was valuable. In these cases, the diagnosis of MS was known, and there were many lesions. However, registration and subtraction helped to identify subtle changes in the presence of other findings. In cases where the finding has undergone dramatic change, alignment is not really needed to diagnose the change, but can still be valuable in assuring that there are no additional subtle findings.

Important advances over prior studies presented in this study are (1) the inclusion of more than a single type of image (CT or T2 MRI); (2) the inclusion of all body parts, not just the head; (3) selection of cases based on actual practice, not ones that were expected to benefit most from alignment; and (4) collection of information about the frequency of prior comparison examinations, allowing estimation of impact on a total radiology practice.

Magnetic resonance images represent a significantly greater challenge than CT because the plane of scanning is much more flexible, and because the contrast properties are very different. Since the time of the report by Schellingerout et al.[10], multidetector CT has made sagittal and coronal reformations much more common, partly negating this difference. However, it can make logistics more difficult, as one CT series can expand to several more, depending on how many volumes are created. An example is creating an angled volume for each disk space. Such complexities will require more thought if automated registration is to be implemented in routine practice.

We found that body imaging, and torso imaging in particular, had a much lower success rate than rigid structures like the head. This is not surprising—one cannot expect to align “fluid” structures using a rigid body assumption. It is possible that a hybrid algorithm that gets an approximate registration, and then refines and registers a region-of-interest defined by a radiologist (perhaps based on an ROI from the prior examination) could increase the effectiveness of registration for images of nonrigid body structures. It is also debatable whether we should have included those cases that we manually aligned because the primary clinical target would be fully automated alignment. Given the low failure rate, this would not have significantly changed the results because most of the failures were in cases where there was low examination similarity.

It may be surprising that in some cases, there was greater agreement for dissimilar examinations than for similar exams whether registered or not (e.g., from Table 1, dissimilar neuro exams had higher agreement than similar neuro exams for both registered and unregistered categories). This may reflect the underlying disease—that certain diseases are easily characterized and so there is low adherence to a standardized protocol. Hence, one should not conclude that intentionally scanning in different fashions improves agreement. A separate study which controls for disease is needed to evaluate that question. The important point is that within each group, (similar or dissimilar) registration showed at least a trend to be advantageous.

CONCLUSIONS

Automated image registration can be performed on clinical images. The results of this study suggest that it may produce small but significant gains in decision time, and better observer agreement for CT and MR studies. However, we note that this study was for a single department, and validation in a larger study including other practices is necessary to validate this conclusion. The benefits are likely greater when similar scanning techniques are used.

ACKNOWLEDGEMENT

This work was presented at the 5th R&D Symposium at the 2005 SCAR meeting with the support of the SCAR R&D committee.

APPENDIX A

The Registration Algorithm Parameters

To improve speed and avoid local minima, registration algorithms often begin with low resolution images to get an approximately correct answer, and then use that to move on to a higher resolution image to improve that estimate until one finally uses the original high resolution image for the final estimate of the relationship between the two data sets. In our case, because the *X* and *Y* dimensions were much smaller, we “shrank” the image eight times more in the *X* and *Y* directions than in the *Z* direction.

There are many options available when performing a rigid alignment of one data set with another. One decision is how many types of motion you will allow to fit one data set onto the other. For instance, one could allow only translation along the three axis (*X*, *Y*, and *Z*). If one also allowed rotation about the three axes, there would be a total of six degrees of freedom. We included these six, plus a single scaling factor for a total of seven degrees of freedom.

Once the geometric relationship between two data sets is determined, it is necessary to "recut" one data set to map onto the other. To do this, it is necessary to create some formula for creating the values between the grid values in the actual image to create the new aligned image. One could simply take the voxel closest to the computed location (known as nearest-neighbor interpolation) but the result is often a jagged appearing output. B-spline is a mathematical technique for creating intermediate values between known values that appears more pleasing.

REFERENCES

1. Chen CT: Radiologic image registration: old skills and new tools. Acad Radiol 10:239–241, 2003

2. Studholme C, Hill DLG, Hawkes DJ: Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures. Med Phys 24:25–35, 1997

3. Woods RP, Mazziotta JC, Cherry SR: MRI-PET registration with automated algorithm. J Comput Assist Tomogr 17:536–546, 1993

4. Collignon A, Maes F, Delaere D, Vandermeulen D, Suetens P, Marchal G: Automated multi-modality image registration based on information theory. In: Bizais Y CB, Di Paola R Eds. Information Processing in Medical Imaging. Dodrecht, The Netherlands: Kluwer, 1995, pp 263–274

5. Hajnal JV, Saeed N, Oatridge A, et al: Detection of subtle brain changes using subvoxel registration and subtraction of serial MR images. J Comput Assist Tomogr 19:677–691, 1995

6. Hill DL, Batchelor PG, Holden M, Hawkes DJ: Medical image registration. Phys Med Biol 46:R1–45, 2001

7. Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P: Multimodality image registration by maximization of mutual information. IEEE Trans Med Imaging 16:187–198, 1997

8. Pelizzari C, Chen G, Spelbring D, Weichselbaum R, Chen C: Accurate three-dimensional registration of CT, PET, and/or MR images of the brain. J Comput Assist Tomogr 13:20–26, 1989

9. Pluim JP, Maintz JA, Viergever MA: Mutual-information-based registration of medical images: a survey. IEEE Trans Med Imaging 22:986–1004, 2003

10. Schellingerhout D, Lev MH, Bagga RJ, et al: Coregistration of head CT comparison studies: assessment of clinical utility. Acad Radiol 10:242–248, 2003

11. Lau PCP, Erickson BJ, Bartholmai B, Knudsen JM, Maus TP: Image Registration: Useful or useless in comparison studies in Body and Neuro MR and CT in PACS environment? In: Langer Ea Ed. Proc SCAR. Vancouver, BC: SCAR, 2004, p 21

12. Mattes D, Haynor D, Vesselle J, Lewellen T, Eubank W: Nonrigid multimodality image registration. In: Sonka M, Hanson K Eds. SPIE. San Diego: SPIE Press, 2001

13. Shrout PE, Fleiss JL: Intraclass correlations: uses in assessing rater reliability. Psychol Bull 86:420–428, 1979

14. Tan IL, van Schijndel RA, Fazekas F, et al: Image registration and subtraction to detect active T(2) lesions in MS: an interobserver study. J Neurol 249:767–773, 2002